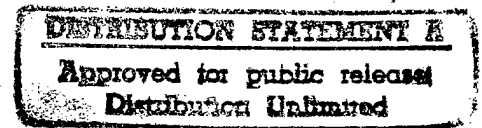


# Analysing Stable Time Series

Robert J. Adler, Raisa E. Feldman, and Colin Gallagher <sup>1</sup>

1997

Abstract



We describe how to take a stable, ARMA, time series through the various stages of model identification, parameter estimation, and diagnostic checking, and accompany the discussion with a goodly number of large scale simulations that show which methods do and do not work, and where some of the pitfalls and problems associated with stable time series modelling lie.

## 1. Introduction

There are three major stages in the now standard "Box-Jenkins" time series modelling techniques for Gaussian time series: Model identification, parameter estimation, and diagnostic checking.

In many ways, the techniques behind these three stages really only involve two bags of tricks, since most diagnostic checks rely on testing whether or not the fitted residuals, after parameter estimation, behave like a white noise sequence. This, of course, is tantamount to identifying a model for the residuals, and so takes us, more or less, back to stage one.

In this paper we will concentrate on a variety of issues related to ARMA model identification in the stable setting. The paper by Calder and Davis, in this volume, [CD], describes the parameter estimation problem, so that the two papers, together, should give a good overview of the overall ARMA problem and be of some assistance to a practitioner who wishes to analyse a particular series. We shall also have something to say about parameter estimation, for one specific technique.

There are no new theorems in this paper, or even really new ways of thinking about things. Rather, we have tried to collect, in one place, a number of results that are rather widely scattered, and to investigate their practical efficiency on replicates of synthetic data. Some of the results are somewhat surprising, and some more than a little worrying. Many beg further, and deeper, theoretical investigation.

---

<sup>1</sup>Research supported in part by Office of Naval Research, grants N00014-94-1-0191, N00014-93-1-0800 and N00014-96-1-0739

DTIC QUALITY INSPECTED 2

19980209 062

The bottom line will be that while, in principle, the standard Gaussian Box-Jenkins techniques [BJ], [BD] do carry over to the stable setting, in practice a great deal of care needs to be exercised.

Results in a similar vein can also be found in the paper [R2] in this volume, as well as [R1] and [FR]. These papers treat real as well as synthetic data, and general heavy tailed rather than purely stable series.

Finally, before we start, we should determine precisely what we mean by the various stable parameters by defining a stable distribution with parameters  $(\alpha, \beta, \sigma, \mu)$  as usual via its characteristic functions, as follows:

$$E(e^{itZ}) = \begin{cases} \exp\{-\sigma^\alpha |t|^\alpha (1 - i\beta(\text{sign } t) \tan \frac{\pi\alpha}{2}) + i\mu t\} & \text{if } \alpha \neq 1, \\ \exp\{-\sigma |t| \left(1 + \frac{2i\beta}{\pi}(\text{sign } t) \ln |t|\right) + i\mu t\} & \text{if } \alpha = 1, \end{cases} \quad (1.1)$$

where  $0 < \alpha \leq 2$ ,  $\sigma \geq 0$ ,  $-1 \leq \beta \leq 1$ , and  $\mu \in \mathbb{R}$ . We shall denote such a distribution by writing  $Z \sim S_\alpha(\sigma, \beta, \mu)$ .

## 2. Preliminary data analysis – Is it stable?

The first question that must be broached is whether or not our data is “heavy tailed”, in some general sense, and, if so, whether or not it is stable. We shall not be interested in the possibility of heavy tailed, but non-stable data, for a number of reasons:

1. If the data is in the domain of attraction (cf. [FE]) of a stable distribution, then, in general, large sample techniques are identical to those for the purely stable situation.
2. In the domain of attraction case, the difference between a stable and non-stable model lies in the central region of the distribution. If one is using stable or other heavy tailed techniques, this is generally not the region of interest.
3. In the case of heavy tails, *not* in a stable domain of attraction, there are comparatively few reliable techniques around (see [R1, FR] for further details and discussion).

We shall also make one significant simplification throughout this paper: We shall virtually always work with examples in which, in terms of (1.1),  $\beta = \mu = 0$ ; i.e. with centered and symmetric variables. This simplification is common in most of the theory that we quote, although, unfortunately, it is not always justified in practice. However, we doubt that it has much *qualitative* effect on the phenomena we shall look at. That this is definitely the case in some situations is born out by [KN2].

## 2.1 Graphing the series.

The simplest, most obvious, and often most powerful, techniques for detecting stable data are also, unfortunately, ones with very little theory behind them. They start, with a visual inspection of the data, in a search for highly “inhomogeneous” data, in the (non-technical) sense that one, or a few observations, dominate the rest. This is generally so notable, that on graphing the time series, most of the data is “squeezed” onto the horizontal axis by the automatic scaling of the plotting routine.

An example of this is given in Figure 1, where plots of four stable time series are given. Each follows the AR(1) model

$$X_t - .5X_{t-1} = Z_t, \quad (2.1)$$

where the  $\{Z_t\}$  are symmetric i.i.d. stables with scaling parameter  $\sigma = 1$ . It is obvious that all three of the stable cases are qualitatively different to the final, Gaussian, case.

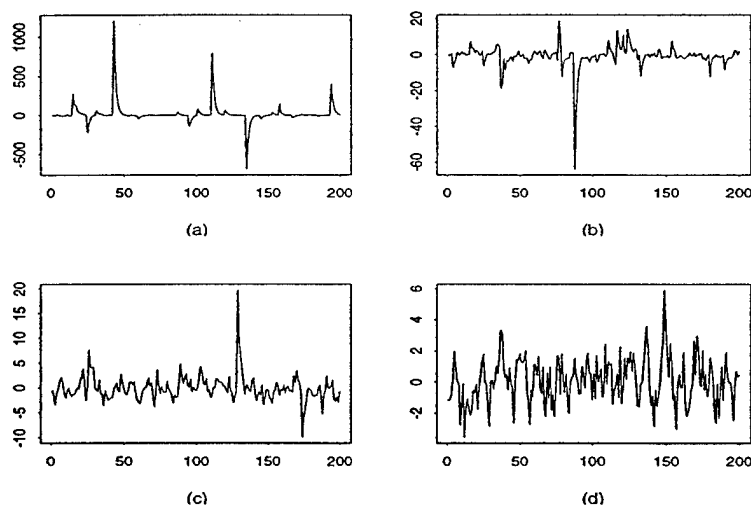


Figure 1: Four AR(1), stable, time series, with increasing values of  $\alpha$ .  
(a)  $\alpha = 0.6$ , (b)  $\alpha = 1.2$ , (c)  $\alpha = 1.8$ , (d)  $\alpha = 2.0$

Of course, in each of these cases, it would be hard to distinguish on a graphical basis between a purely stable series and a Gaussian series with the occasional outlier (cf. the examples in [MI]). This requires looking more carefully at further distributional information.

## 2.2 The histogram.

Essentially the same information as is obtained by graphing the series can be gleaned from the histogram of the data. What is lost in the histogram is, of

course, the temporal structure of the data, but what is more obvious is the presence or absence of symmetry.

Figure 2 shows a histogram from data generated by the same model as in (2.1), but now only for two cases,  $\alpha = 2$  (Gaussian) and  $\alpha = 1.6$ , and for series of length 1,000. Two factors should be noted here. The first is that, despite the fact that the sample size is quite large, the Gaussian case is much further from the traditional bell curve than one would expect with i.i.d. data. But, this is correlated data, so that the laws of large numbers take longer to come into play.

The second is a repetition of the phenomenon mentioned above, about automatic scaling “spoiling” the graph. In (b) a few outliers are so large that the entire histogram is squeezed into a few bars in the center. When the largest and the smallest 5% of the data is truncated, as in (c), the shape of the graph changes dramatically.

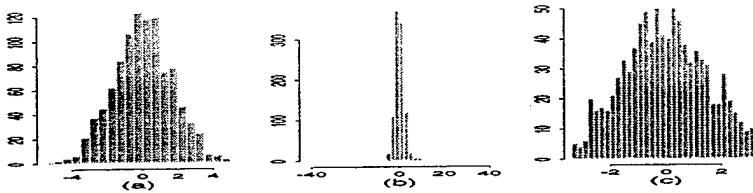


Figure 2: Histograms of (a) Gaussian, (b) Stable,  $\alpha = 1.6$ , and (c) Truncated stable time series.

### 2.3 The “converging variance” test.

One of the oldest tests for determining whether data has infinite variance is the trick of plotting the sample variance  $S_n^2$ , based on the first  $n$  observations, as a function of  $n$ . If the data comes from population with finite variance,  $S_n^2$  should converge to a finite value. Otherwise, it should diverge as  $n$  grows, and the graph typically shows large jumps.

Although this test was originally designed for i.i.d. data, it also works well for correlated data, as long as the order of the observations is first randomised, so as to destroy dependencies that might lead to trends and jumps with other explanations.

Figure 3 contains graphs for this test for two stable ( $\alpha = 0.8, 1.6$ ), Gaussian, and “t” (with 4 degrees of freedom) processes. (By the last we mean an ARMA process in which the innovations  $Z_t$  have a Student t distribution. This is an interesting case, since it gives a distribution with much heavier than Gaussian tails, but in the Gaussian domain of attraction.) For variety, we took the ARMA(1,1) model  $X_t - 0.3X_{t-1} = Z_t + .5Z_{t-1}$ , with series of length 1,000.

The divergence of  $S_n^2$ , as  $n$  grows, and the irregularity of the graphs in the two stable cases are very marked.

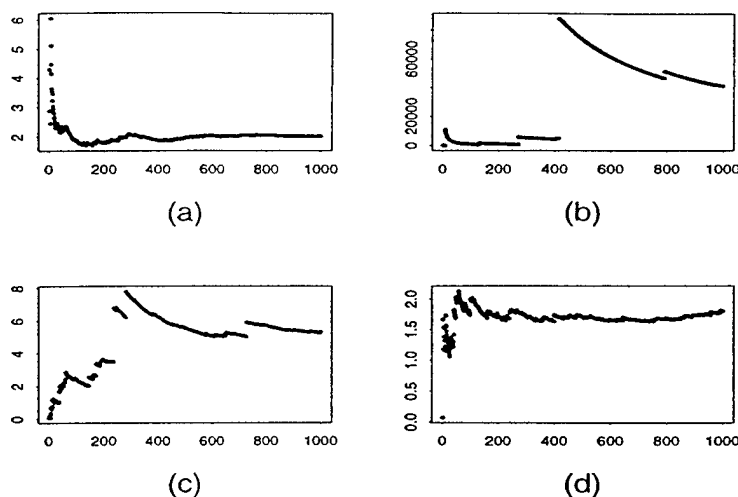


Figure 3: Cumulative variance plots for (a) Gaussian, (b)  $\alpha = 0.8$ , (c)  $\alpha = 1.6$  and (d) Student t ARMA(1,1) processes.

## 2.4 Preliminary estimation of the stable parameters.

One last, and natural, step before entering the time series arena proper, is the estimation of the various stable parameters.

There are a number of techniques available for estimating tail decay, most of which are built around the so-called “Hill estimator”, and many of which, while based on sound theory, turn out to be far from satisfactory in practice. (cf. [R1, R2] or [PDM], in this volume, for details.)

This, indeed, is one of the reasons for being prepared to assume a specifically stable model, rather than one generically heavy tailed. For in the stable case there exists an excellent estimator of the parameters, due to Hu McCulloch [MC]. This estimator, which is based in essence on fitting tabulated quantiles of stable distributions, works for  $\alpha \in [0.6, 2.0]$  and  $\beta \in [-1, 1]$  (which covers most of the cases met in practice) and all values of the other parameters. The estimator of the location parameter however, can be inaccurate near  $\alpha = 1$ .

The McCulloch estimator was originally designed for, and indeed works best on, i.i.d. data. Nevertheless, some initial information on the parameters, especially  $\alpha$ , is generally required for model identification, so that one has no choice but to work with the time series data.

A typical example of the precision of the McCulloch estimator of  $\alpha$  is given in Table 1, in which the results for estimating  $\alpha$  from 1,000 observations from the MA(2) model  $X_t = Z_t + .5Z_{t-1} - .3Z_{t-2}$  are presented for various  $\alpha$ . The accuracy of the estimator is, we believe, truly impressive.

For each iteration we estimate  $\alpha$  from the simulated innovations, and then

from the time series. For the final column, we estimated the MA parameters using Whittle's estimator, described in Section 4, and computed the residuals. We then estimated  $\alpha$  a third time using these residuals, expecting (incorrectly) that this estimation would be better than from the raw time series. We repeated this process 10,000 times. Clearly the estimates obtained from the innovations are the best, but, perhaps rather surprisingly, McCulloch's technique seems to work better when applied to the original time series rather than to the residuals.

Estimation of alpha, MA(2)			
$\alpha$	Innovations	Time series	Residuals
.6	0.633 (0.0397)	0.665 (0.0796)	0.679 (0.1010)
.8	0.804 (0.0363)	0.811 (0.0524)	0.822 (0.0617)
1	1.002 (0.0439)	1.007 (0.0593)	1.014 (0.0666)
1.2	1.203 (0.0506)	1.206 (0.0648)	1.209 (0.0724)
1.4	1.402 (0.0582)	1.405 (0.0690)	1.407 (0.0769)
1.6	1.606 (0.0711)	1.608 (0.0791)	1.610 (0.0888)
1.8	1.808 (0.0935)	1.810 (0.0977)	1.812 (0.1023)
2	1.953 (0.0653)	1.954 (0.0659)	1.952 (0.0683)

Table 1: *Mean and standard deviation (in parentheses) of 10,000 estimates of alpha using simulated innovation sequence, corresponding time series and estimated residuals.*

Interestingly, however, McCulloch's estimator does not seem to work as well for AR processes as it does for pure moving averages, at least for small values of  $\alpha$ , the main problem being in the substantially increased sample variance, rather than in the bias.

There are a number of possible explanations for this, although we are not certain which, if any, is real. Two candidates for consideration are:

- (i) It may simply be due to divergences, numerical and other, as  $\alpha$  becomes close to the region where the estimator is not supposed to work.
- (ii) In the Gaussian case, the correlation structure is much stronger in the AR(2) model than in the MA(2), and this should affect estimator variance. Translating "correlation" to "dependence" in the stable case, may create a similar problem.

However, when estimating  $\alpha$  from the residuals obtained from the Whittle parameter estimates, the simulations indicate that the estimates are superior in the AR case (at least for  $\alpha \geq 1$ ).

Table 2 gives the result of a similar study for the model  $X_t - .8X_{t-1} + .7X_{t-2} = Z_t$ .

Note that these positive results become less than satisfactory when one leaves the permissible parameter region for the estimator, and in the region of  $\alpha = 1$ . We simulated a sample of length 1000 from a of symmetric Cauchy distribution ( $\alpha = 1$ ,  $\mu = 0$ ) and used McCulloch's estimator to estimate the

location parameter  $\mu$ . The mean of 50 of such independent estimates of location was -6.17.

Similarly, estimates of  $\alpha$  for values of  $\alpha < 0.6$  also give poor and misleading results. For example, the average of 100 estimates of  $\alpha$  based on independent samples of 1000 was 5.23 for  $\alpha = 0.3$ , and 0.95 for  $\alpha = 0.5$ .

Estimation of alpha, AR(2)			
$\alpha$	Time series		Residuals
.6	0.836	(0.7122)	0.790 (0.6434)
.7	0.758	(0.2976)	0.740 (0.2822)
.8	0.843	(0.1040)	0.818 (0.0702)
1	1.033	(0.0946)	1.009 (0.0561)
1.2	1.213	(0.0975)	1.207 (0.0544)
1.4	1.408	(0.1007)	1.405 (0.0588)
1.6	1.612	(0.1044)	1.608 (0.0719)
1.8	1.813	(0.1079)	1.811 (0.0941)
2	1.954	(0.0670)	1.954 (0.0655)

Table 2: Mean and standard deviation (in parentheses) of 10,000 estimates of alpha using simulated time series and estimated residuals.

### 3. Model Identification

The first step when fitting data  $\{X_1, X_2, \dots, X_n\}$  into a linear ARMA( $p, q$ ) time series model

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad (3.1)$$

with i.i.d. innovations  $\{Z_t\}$ , normal or stable, is the identification of the lag parameters  $p$  and  $q$ .

In the Gaussian case model identification techniques are based on analysis of the sample autocorrelation function (ACF)

$$\hat{\rho}(h) = \sum_{t=1}^{n-h} X_t X_{t+h} / \sum_{t=1}^n X_t^2, \quad h = 1, 2, \dots, \quad (3.2)$$

or its mean-corrected version,

$$\bar{\rho}(h) = \sum_{t=1}^{n-h} (X_t - \bar{X})(X_{t+h} - \bar{X}) / \sum_{t=1}^n (X_t - \bar{X})^2, \quad (3.3)$$

where  $\bar{X} = n^{-1}(X_1 + \dots + X_n)$ , and of the equally familiar sample partial autocorrelation function (PACF), which, to make some details below easier to follow, we define in detail:

Consider the AR( $p$ ) model

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t, \quad (3.4)$$

$1 - \phi_1 z - \dots - \phi_p z^p \neq 0$ ,  $|z| \leq 1$ . Let  $R_p = [R_{i,j}]_{i,j=1}^p = [\rho(i-j)]_{i,j=1}^p$  be the  $p \times p$  matrix of ACFs computed under a finite variance assumption,  $\rho = (\rho(1), \dots, \rho(p))'$ ,  $\phi = (\phi_1, \dots, \phi_p)'$ . Then the Yule-Walker matrix equation is

$$R_p \phi = \rho, \quad (3.5)$$

and the Yule-Walker estimate of  $\phi$  is then defined as solution of

$$\hat{R}_p \hat{\phi} = \hat{\rho}, \quad (3.6)$$

where  $\hat{R}_p = [\hat{\rho}(i-j)]_{i,j=1}^p$  and  $\hat{\rho} = (\hat{\rho}(1), \dots, \hat{\rho}(p))'$ .

To define the PACF in the AR case, we consider vectors  $\phi_m^* = (\phi_1^*, \dots, \phi_m^*)$ , where  $\phi_i^* = \phi_i$  for  $i \leq p$  and  $\phi_i^* = 0$  when  $i > p$ . For this vector we write Yule-Walker equation (3.5) as  $R_m \phi_m^* = \rho_m$ , where  $\rho_m = (\rho(1), \dots, \rho(m))$ . The PACF at lag  $m$ ,  $\phi_{mm}$ , is then defined as the  $m$ -th component of the vector  $\phi_m^* = R_m^{-1} \rho$ . Similarly, the sample PACF function at lag  $m$ ,  $\hat{\phi}_{mm}$ , is defined as the  $m$ -th component of the vector

$$\hat{\phi}_m^* = \hat{R}_m^{-1} \hat{\rho}. \quad (3.7)$$

In the heavy tailed case the ACF and PACF do not exist, but we can still use equations (3.2) – (3.7) to define their sample equivalents.

In the general ARMA case, the PACF at lag  $h$  is defined as the sample correlation between the residuals of  $X_{t+h}$  and  $X_t$  after linear regression (under a finite variance assumption) on  $X_{t+1}, X_{t+2}, \dots, X_{t+h-1}$ .

We shall use the notation  $\tilde{\phi}_{kk}$  for the PACF's when the centered ACFs  $\tilde{\rho}$ 's of (3.3) are used in the Yule-Walker equation (3.6) instead of the non-centered  $\hat{\rho}$ 's of (3.2). For most of the simulations we shall prefer to use the centered variables, and so will need to assume that  $\alpha > 1$ .

We shall see, basically via simulation, that both the ACF and PACF provide excellent tools for studying stable time series. It is perhaps rather surprising that although second moments are infinite in the stable case, the tools that we are used to from the Gaussian case are still available, albeit with some modifications.

The following subsection sets up the theory underlying the use of the ACF and PACF. We then provide some tables for hypothesis testing, followed by two sections on numerics. The final subsection looks at the use of the Akaike Information Criterion in the stable setting.

### 3.1 The basic theory.

The theoretical basis for the usage of the sample ACF for the identification of the order  $q$  of a MA( $q$ ) time series is the following fundamental result of [DR2] (we follow here [BD] Theorem 13.3.1):

**Theorem 3.1** Let  $\{Z_t\}$  be an iid symmetric sequence of  $\alpha$ -stable random variables and let  $\{X_t\}$  be the strictly stationary process,

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad (3.8)$$

where

$$\sum_{j=-\infty}^{\infty} |j| |\psi_j|^\delta < \infty \quad \text{for some } \delta \in (0, \alpha) \cap [0, 1].$$

Define for such a process an analogue of the ACF function, namely

$$\rho(h) = \sum_j \psi_j \psi_{j+h} / \sum_j \psi_j^2, \quad h = 1, 2, \dots \quad (3.9)$$

Then, for any positive integer  $h$ ,

$$(n/\ln n)^{1/\alpha} (\hat{\rho}(1) - \rho(1), \dots, \hat{\rho}(h) - \rho(h))' \Rightarrow (Y_1, \dots, Y_h)', \quad (3.10)$$

where

$$Y_k = \sum_{j=1}^{\infty} (\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)) S_j / S_0. \quad (3.11)$$

Here  $S_0, S_1, \dots$  are independent stable variables;  $S_0$  is positive with  $S_0 \sim S_{\alpha/2}(C_{\alpha/2}^{-2/\alpha}, 1, 0)$ , and the  $S_j \sim S_{\alpha}(C_{\alpha}^{-1/\alpha}, 0, 0)$  where

$$C_{\alpha} = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\frac{\pi\alpha}{2})} & \text{if } \alpha \neq 1 \\ \frac{2}{\pi} & \text{if } \alpha = 1. \end{cases} \quad (3.12)$$

The marginal distribution of each  $Y_k$  is somewhat simpler, and we have

$$Y_k = \left( \sum_{j=1}^{\infty} |\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)|^\alpha \right)^{1/\alpha} U/V, \quad (3.13)$$

where  $V$  and  $U$  are independent stable random variables with the same distributions as  $S_0$  and  $S_1$  in (3.11).

When  $\alpha > 1$  then (3.10) is also true when  $\hat{\rho}(h)$  is replaced by its mean-corrected version,  $\bar{\rho}(h)$ .

We consider now an example of this theorem in practice. Let  $X_t$  be the symmetric stable MA( $q$ ) process,

$$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}. \quad (3.14)$$

Then the above theorem implies that

$$(n/\ln n)^{1/\alpha} (\bar{\rho}(h) - \rho(h)) \Rightarrow \left( 1 + 2 \sum_{j=1}^q |\rho(j)|^\alpha \right)^{1/\alpha} U/V, \quad h > q, \quad (3.15)$$

where the right hand side reduces to  $U/V$  if  $q = 0$ .

Thus one can use the above results to plot confidence intervals for the ACF function and identify the parameter  $q$ , once the distribution of  $U/V$  is known.

A similar result also holds for the distribution of the PACF. Since  $\hat{\rho} \xrightarrow{P} \rho$  and  $\hat{R}_p \xrightarrow{P} R_p$ , the consistency of the Yule-Walker estimates follows. In fact, the mean value theorem gives

$$\hat{\phi} - \phi = D(\hat{\rho} - \rho) + o_p(\hat{\rho} - \rho),$$

where  $D$  is the  $p \times p$  matrix of partial derivatives of vector function  $\psi(\mathbf{z}) := R_p(\mathbf{z})^{-1}\mathbf{z}$ . Here  $R_p(\mathbf{z}) = [z_{|i-j|}]_{i,j=1}^p$ ,  $z_0 \equiv 1$  and  $\phi = \psi(\rho)$ . Theorem 3.1 then yields that

$$(n/\ln n)^{1/\alpha}(\hat{\phi} - \phi) \Rightarrow D(Y_1, \dots, Y_p)' \quad (3.16)$$

where the vector  $(Y_1, \dots, Y_p)$  has distribution described by (3.11) - (3.13).

The limiting distribution of the PACF's is now given by (3.16) and (3.11) - (3.13), and so is, in general, rather complicated. However, when  $p = 0$ , the right hand side of (3.16) reduces to  $U/V$ , which is the same limit as for the ACF of white noise. Since in the null hypothesis case this is what is required to test which of the  $\phi_{mm}$  are zero, we are in the fortunate situation of being able to use the same distribution twice.

### 3.2 Some important quantiles.

In practice, the distribution of  $U/V$  cannot be computed theoretically, but only via simulation or numerical integration of the joint density of the vector  $(U, V)$  over an appropriate region.

Figure 4 gives the density of  $U/V$  for two values of  $\alpha$ . Note the high tails of the distribution, which are high even in relation to stable distributions.

Table 3 contains the 97.5% quantiles of  $U/V$  (the distribution of  $U/V$  is symmetric) which for  $\alpha < 2$  were found via simulation of 500,000 values of  $U/V$  using a corrected version of the S-plus routine for generating stable random variables. (cf. [ST] p43. The S-plus routine does not quite deliver what you might expect in the asymmetric case!) Our value for  $\alpha = 1$  coincides with the value quoted in [BD], p541, and found via numerical integration.

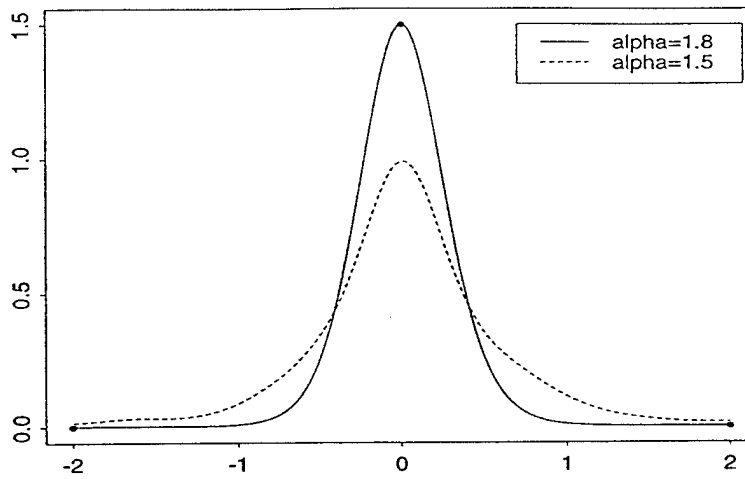


Figure 4: Kernel density estimates of  $U/V$  for  $x \in [-2, 2]$  and  $\alpha = 1.5$  and  $1.8$ .

Table 3: 97.5% quantiles of $U/V$			
$\alpha$	97.5% quantile	$\alpha$	97.5% quantile
0.3	9.338e+04	1.3	3.865e+00
0.4	4.625e+03	1.4	2.814e+00
0.5	7.375e+02	1.5	2.059e+00
0.6	1.986e+02	1.6	1.516e+00
0.7	7.745e+01	1.7	1.096e+00
0.8	3.710e+01	1.75	9.280e-01
0.9	2.072e+01	1.8	7.637e-01
1.0	1.240e+01	1.9	4.765e-01
1.1	8.088e+00	2.0	1.960e+00
1.2	5.532e+00		

### 3.3 Estimating the lag parameters via the ACF and PACF.

To see how useful the above results are, and to compare them to an attempt to estimate the parameters  $p$  and  $q$  assuming normality, we conducted the following, rather illuminating, double blind study.

We simulated 100 time series of length  $n = 1000$  with symmetric stable innovations with  $\alpha = 1.2$ . The models were selected at random from the following five models:

$$\begin{aligned}
 X_t &= Z_t, \\
 X_t &= Z_t - .7Z_{t-1}, \\
 X_t &= Z_t - .7Z_{t-1} + .8Z_{t-2},
 \end{aligned}$$

$$\begin{aligned} X_t - .7X_{t-1} &= Z_t, \\ X_t - .7X_{t-1} + .8X_{t-2} &= Z_t. \end{aligned}$$

We plotted the ACF and PACF for each of these series, and then used these plots to try to identify the true model using the standard time series technique of looking at the plots and seeing how they behave relative to the 95% confidence intervals under a white noise null hypothesis. Since we did not want to assume  $\alpha$  to be known, we used confidence intervals corresponding to the Gaussian ( $\alpha = 2$ ) and Cauchy ( $\alpha = 1$ ) cases. (In fact, the confidence intervals for the true  $\alpha = 1.2$  and Cauchy cases were indistinguishable to the human eye. The Gaussian intervals, however, were about 28% shorter.) The conclusions were then compared against the true models.

The procedure showed 31% error when Gaussian limits were used and 17% error using Cauchy limits. Although the error rate clearly depends on experience of a person doing identification, it is clear from this study that using stable limits for heavy-tailed data reduces the error rate significantly.

### 3.4 On asymptotics or “a funny thing happened on the way to $\infty$ ”.

A fact often touted by stable time series theorists as a compensation for the difficulties generally associated with stable rather than Gaussian analysis is that the rate of convergence of  $\bar{\rho}(h) - \rho(h)$  to zero is of the order  $O([n/\ln n]^{-1/\alpha}) = o(n^{-1/\beta})$  for all  $\beta > \alpha$ , which is considerably faster than in Gaussian case when the rate is on the order of  $O(n^{-1/2})$ .

However, despite this comforting fact, there are some other, rather problematic, phenomena associated with this convergence, since the rate of convergence of the *distribution* of  $\bar{\rho}(h) - \rho(h)$  to the limiting distribution is actually very slow.

Before we mention some theory, consider Table 4, which indicates how fast (or slow) the distribution of  $\bar{\rho}(1)$  converges to the theoretical one for the white noise model. For values of  $\alpha = 1.4, 1.7, 1.75$ , and  $1.8$  and sample size  $n$  we computed 10,000 coefficients  $\bar{\rho}(1)$  from independent white noise sequences with corresponding values of  $\alpha$  and  $n$  and checked the percentage of times the coefficient was *not* within the nominal 95% confidence interval. (Of course, this should be 5%.) For determining confidence intervals we used three different distributions: a stable distribution with the correct  $\alpha$  (as described above), a Cauchy distribution and Gaussian distribution. (i.e. we used either the correct value of  $\alpha$ , or behaved as if  $\alpha = 1$  or  $2$ .) The results show that when the correct stable distribution is used, the convergence to theoretical 5% error is very slow, and that for “small” sample sizes of the order of 1000 the Cauchy based limits actually give the best results.

The main reason behind this phenomenon seems to be the slow convergence of the distributions of stable averages to their limiting distributions, a fact that has a well documented history. (cf. [CH, HA, HW, JM] although none of these quite treats our setting.)

Table 4: Percent error for 95% confidence interval of  $\tilde{\rho}(1)$ , white noise

$\alpha$	$n = 1000$	$n = 10,000$	$n = 500,000$	$n = 1,000,000$
(i) $\alpha$ -stable limits				
1.4	1.43	2.28	3.08	3.27
1.7	4.37	4.58	4.68	3.67
1.75	6.53	5.35	5.40	5.60
1.8	10.55	8.65	7.44	6.29
(ii) Cauchy limits				
1.4	1.24	6.62	32.27	40.10
1.7	0.80	15.87	69.87	73.95
1.75	0.70	17.19	73.07	79.65
1.8	0.74	19.79	76.04	80.69
(iii) Gaussian limits				
1.4	2.91	2.17	0.85	0.89
1.7	3.46	3.36	2.14	1.35
1.75	3.56	3.23	2.72	2.52
1.8	4.22	3.68	2.90	2.41

From the practical point of view, this phenomenon shows up in an interesting way in the numerics. Figure 5 shows the values of the upper limits of the confidence intervals for  $(\tilde{\rho}(1) - \rho(1))$  for various  $\alpha$  and  $n$ , based on the limiting distribution of Theorem 1 and a white noise time series. Note how, for "small"  $n$ , the confidence interval shrinks to a point as  $\alpha \rightarrow 2$ .

We are not certain of the reason for this. One possibility is that for small  $n$  the limiting distribution of  $U/V$  is not appropriate. What seems more reasonable, however, is that the norming constants used to obtain the limiting distribution, while asymptotically correct, are too large (in an  $\alpha$  dependent fashion) for small and intermediate values of  $n$ .

It is not totally clear how to get around this problem without using, perhaps, something like a bootstrap. However, there is growing evidence that in the stable situation bootstrapping is also problematic (cf. [LPPR] in this volume.)

One possible approach comes out of Table 5, which explores the  $n = 1000$  case for different  $\alpha$ 's. Our model is again white noise, and we record the percentage of times when the sample coefficient  $\tilde{\rho}(1)$  is outside the 95% confidence interval; i.e. the percentage of wrong identifications of the model when the identification procedure is done by computer and is based on the value of  $\tilde{\rho}(1)$  only. The confidence levels were computed based on the true  $\alpha$ -stable distribution, as well as Cauchy and Gaussian distributions. The number of simulations for each case was  $m = 10,000$ .

It seems that the Gaussian limits perform the most poorly, at least for  $\alpha < 1.7$  and that for all  $\alpha$  Cauchy limits not only perform better than the others, but do quite well on an absolute scale. (The error here is never larger

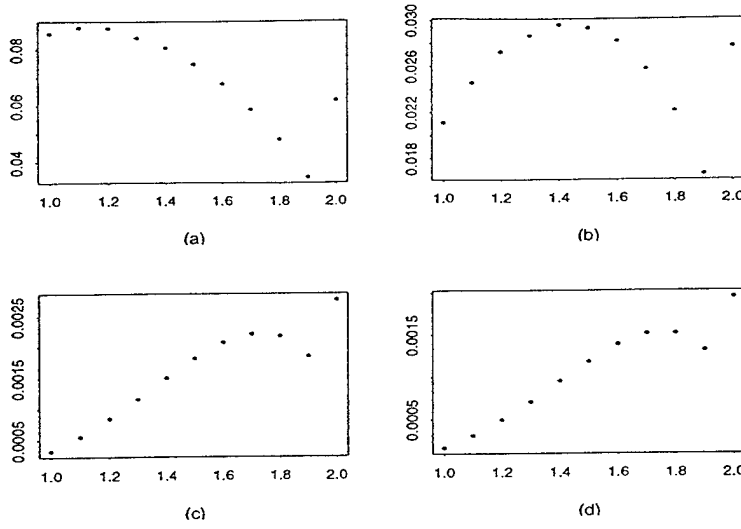


Figure 5: Confidence intervals for  $(\tilde{\rho}(1) - \rho(1))$  for  $1 \leq \alpha \leq 2$  and various values of  $n$ . (a)  $n = 1,000$ , (b)  $n = 5,000$ , (c)  $n = 500,000$ , (d)  $n = 1,000,000$

than 1.28%). Furthermore, for small  $\alpha$  there is no significant benefit in using the confidence limits based on the true stable distribution. However, for large  $\alpha$ , the rate of the rate of convergence of  $\tilde{\rho}(1)$  to its theoretical distribution is so slow that Cauchy or Gaussian distribution should be used. Comparison with Table 4 strengthens this point.

$\alpha$	true $\alpha$ -stable distribution	Cauchy limits	Gaussian limits
1.0	1.12	1.12	2.03
1.1	1.20	1.28	2.32
1.2	1.13	1.22	2.46
1.3	1.27	1.23	2.61
1.4	1.43	1.24	2.91
1.5	1.85	1.18	3.26
1.6	2.55	1.19	3.33
1.7	4.37	0.80	3.46
1.8	10.55	0.74	4.22
1.9	25.63	0.42	4.55
2.0	4.70	0.77	4.70

In Table 6 we continue the theme of using Cauchy based bounds, regardless of the true value of  $\alpha$ . While this clearly gives a conservative test, it turns out that in practice it is not overly so, and the results here illustrate how amazingly well the Cauchy bounds perform in the identification of the MA(1)

model  $X_t = Z_t - .8Z_{t-1}$ . Again, each specific case was run  $m = 10,000$  times. In each run for sample size  $n = 1000$  and given  $\alpha$  we computed  $\bar{\rho}(1), \dots, \bar{\rho}(10)$  and recorded an "error" when  $\bar{\rho}(1)$  was within the confidence limits for white noise or when  $\bar{\rho}(1)$  was outside these limits but one of the coefficients  $\bar{\rho}(2), \dots, \bar{\rho}(10)$  was also outside the limits, so identification as MA(1) would not be called for. The confidence limits were taken to be (i)  $(\ln n/n)^{1/\alpha}(U/V)$  with  $U$  and  $V$  coming from the true distribution, (ii) as in (i) but with  $U$  and  $V$  corresponding to Cauchy distribution and with  $\alpha = 1$ , (iii)  $1.96/\sqrt{n}$ , (iv) according to Bartlett's formula for MA(1) model with true value of  $\rho(1)$ . The whole procedure was performed by computer without human intervention.

It is clear, although perhaps somewhat surprising, that Cauchy limits work the best and that the identification procedure works better for heavy-tailed series than for their finite variance counterparts.

Table 6: Percent of wrong identifications of MA(1) model

$\alpha$	true $\alpha$ -stable	Cauchy	Gaussian	Bartlett
1.0	10.31	10.31	17.36	12.97
1.1	11.61	12.13	20.36	14.81
1.2	11.48	12.08	22.15	15.38
1.3	12.96	12.44	25.10	16.64
1.4	14.89	12.53	27.75	17.80
1.5	20.33	13.80	31.99	19.71
1.6	28.19	13.60	35.52	20.80
1.7	45.59	13.50	40.33	22.32
1.8	68.55	13.85	43.49	23.95
1.9	91.59	14.86	48.39	26.17
2.0	52.59	17.21	52.59	29.73

We close this subsection with a some brief information on the asymptotic behaviour of the PACF, which, not surprisingly, is similar to that of the ACF. Table 7 is a PACF version of Table 5, generated in the same fashion, however with data on  $\tilde{\phi}_{11}$  rather than  $\bar{\rho}(1)$  being tabulated.

Table 7: Percent error for 95% confidence interval of $\phi_{11}$ , $n = 1000$ , WN			
$\alpha$	<i>true <math>\alpha</math>-stable distribution</i>	<i>Cauchy limits</i>	<i>Gaussian limits</i>
1.0	1.07	1.07	1.93
1.1	1.27	1.30	2.39
1.2	1.13	1.25	2.44
1.3	1.43	1.36	2.62
1.4	1.44	1.20	3.11
1.5	1.87	1.30	3.23
1.6	2.45	1.08	3.33
1.7	4.64	0.84	3.86
1.8	10.52	0.70	4.35
1.9	25.93	0.47	4.13
2.0	4.84	0.83	4.84

### 3.5 The AIC criterion.

In the finite variance case the prime criterion for automated model selection recommended by [BD] is the AICC, a modified version of Akaike's AIC criterion. An investigation of the AIC criterion in the infinite variance situation was carried out by [BH] and [KN1].

For an autoregressive model the AIC statistics are defined by

$$AIC(k) = n \ln \hat{\sigma}^2(k) + 2k,$$

where  $n$  is the sample size, and  $\hat{\sigma}^2(k)$  is the estimate of the innovation variance obtained from Yule-Walker estimates for  $k$ -th order autoregressive sequence (cf. (3.6)). Then

$$\hat{p} = \operatorname{argmin}_{k \leq K} AIC(k),$$

where  $K$  is an acceptable upper bound for  $p$ , is the corresponding estimate of the order  $p$ . [KN1] showed that this procedure is consistent for heavy tailed situations.

To see how the AIC criterion works in practice for stable AR series we performed 1000 simulations of the AR(1) model

$$X_t = 0.4X_{t-1} + Z_t, \quad (3.17)$$

for  $\alpha = 0.5, 1.0, 1.75, 2.0$ . The sample size was fixed at  $n = 200$ , which is rather small by stable standards, where larger sample sizes are required than in the Gaussian case.

In each run we assumed that we were looking for the best AR model; i.e. we looked for order  $p$  which minimised the AIC criterion. We then recorded the number of times that the correct order ( $p = 1$ ) was correctly identified.

To check how the AIC criterion works for a MA model, we used the S-plus *arima.mle* routine which defines AIC as  $(-2)$  times the log Gaussian likelihood

plus two times the number of parameters fit (see the definition in [BD], p. 304 or consult the S-plus manual). We performed 1000 runs for model:

$$X_t = Z_t + 0.8Z_{t-1} \quad (3.18)$$

with  $\alpha = 0.5, 1.0, 1.75, 2.0$ ,  $n = 200$ . Again, in each run we assumed that we were looking for the best MA model, and recorded the number of correct identifications. The results of simulations for both models are given in Table 8 and the conclusion is that in both cases the AIC criterion works better the heavier the tails!

The MA case is particularly interesting, since we have no theoretical justification for applying the AIC based on the Gaussian likelihood to heavy-tailed data.

#### 4. Parameter estimation for ARMA models: The Whittle estimator.

Paper [CD] of this volume gives an extensive review of estimation techniques for linear processes with stable innovations. They present convincing evidence to the effect that LAD and MLE estimators are superior, in the heavy tailed setting, to the estimators traditionally used in finite variance time series, such as least square or Whittle (periodogram) estimators. However, for both of these cases, the sampling distribution of the parameter estimates is, at least at the moment, numerically as well as theoretically intractable.

Table 8: Percent of correct model identifications made by AIC.		
$\alpha$	$AR(1) - (3.17)$	$MA(1) - (3.18)$
0.5	96.1	92.5
1.0	89.6	89.0
1.75	78.7	79.7
2.0	73.2	72.5

However, for the Whittle estimator, we have the following result of [MGKA]:

**Theorem 4.1** *Let  $\{X_t\}$  be a causal, invertible,  $\alpha$ -stable,  $ARMA(p, q)$  process, with parameters  $\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)'$ . Let  $C$  denote the space of permissible parameter values: i.e.  $C = \{\beta \in \mathbb{R}^{p+q} : \phi(z) \cdot \theta(z) \neq 0 \text{ for } |z| \leq 1 \text{ and } \phi(\cdot), \theta(\cdot) \text{ have no common zeroes}\}$ . Define the polynomials*

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p, \quad \theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q \quad (4.1)$$

and, for all  $-\pi < \lambda \leq \pi$ , the "power transfer function"

$$g(\lambda; \beta) = \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}. \quad (4.2)$$

Denote the self-normalized periodogram by

$$I_{n,X}(\lambda) = \frac{\left| \sum_{t=1}^n X_t e^{-i\lambda t} \right|^2}{\sum_{t=1}^n X_t^2}, \quad -\pi < \lambda \leq \pi. \quad (4.3)$$

The periodogram, or “Whittle” estimator  $\bar{\beta}_n$  of the true, but unknown, parameter  $\beta_0$ , is found by minimizing

$$\sigma_n^2(\beta) = \frac{2\pi}{n} \sum_j \frac{I_{n,X}(\lambda_j)}{g(\lambda_j; \beta)} \quad (4.4)$$

with respect to  $\beta \in C$ , where the sum is taken over all frequencies  $\lambda_j = 2\pi j/n \in (-\pi, \pi]$ . Then,

$$\left( \frac{n}{\ln n} \right)^{-1/\alpha} (\bar{\beta}_n - \beta_0) \Rightarrow 4\pi W^{-1}(\beta_0) \frac{1}{S_0} \sum_{k=1}^{\infty} S_k b_k, \quad (4.5)$$

where the  $S_0, S_1, \dots$  are as in Theorem 1,  $W^{-1}(\beta_0)$  is the inverse of the matrix

$$W(\beta_0) = \int_{-\pi}^{\pi} \left[ \frac{\partial \ln g(\lambda; \beta_0)}{\partial \beta} \right] \left[ \frac{\partial \ln g(\lambda; \beta_0)}{\partial \beta} \right]^T d\lambda,$$

and

$$b_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ik\lambda} g(\lambda; \beta_0) \frac{\partial g^{-1}(\lambda; \beta_0)}{\partial \beta} d\lambda,$$

where  $g^{-1}$  denotes the reciprocal of  $g$ .

Self-normalization of the periodogram, as in (4.3), is essential for the proof of the above Theorem. However, in practice, in order to find the estimator  $\bar{\beta}_n$  we minimize expression (4.4) when self-normalized periodogram  $I_{n,X}(\lambda)$  is replaced by

$$\hat{I}_{n,X}(\lambda) = \left| n^{-1/\alpha} \sum_{t=1}^n X_t e^{-i\lambda t} \right|^2 \quad (4.6)$$

One of the useful aspects of this result is that the asymptotic sample distribution involved here is closely related to that which arises in the study of the ACF and PACF, so that simple extensions of the numerics required to find confidence intervals there also work here.

To see how close the theoretical asymptotic distribution of Whittle estimator is to the true sample distribution when  $n$  is fixed, we simulated data using the AR(1) model (3.17). We ran three different sample sizes  $n = 200, 1000$  and  $10,000$  and two values of  $\alpha = 1.5, 1.75$ . For comparison we also simulated Gaussian data and used confidence intervals for the MLE, the asymptotic distribution of which coincides with the distribution of Whittle estimates in the finite variance case. We ran  $m = 10,000$  simulations for each particular case.

The results, described in Table 9 and on Figure 6, show that the confidence intervals based on the limiting distribution (4.5) work very well even for the

relatively small sample size of  $n = 200$ . The entries in Table 9 are the percent of estimates  $\bar{\phi}_n \equiv \{\bar{\phi}_1\}_n$  which fall into the 95% confidence interval centered around the true parameter value. Although, the numbers in the Gaussian row are the closest to theoretical 95%, the results for  $\alpha = 1.75$  are impressive and unexpected, given the slow convergence rate of the distributions of the sample ACF and PACF. Figure 6 gives histograms for the Whittle estimate  $\{\bar{\phi}_1\}_n$  for  $\alpha = 1.75$  and the three different sample sizes used. The vertical bars represent 95% confidence levels.

Table 9: Percent of $\{\bar{\phi}_1\}_n$ 's within 95% confidence interval; model (3.17)			
$\alpha$	sample size $n=200$	sample size $n=1000$	sample size $n=10000$
1.5	99.35	98.48	96.41
1.75	92.71	93.76	95.08
2	94.77	94.55	94.89

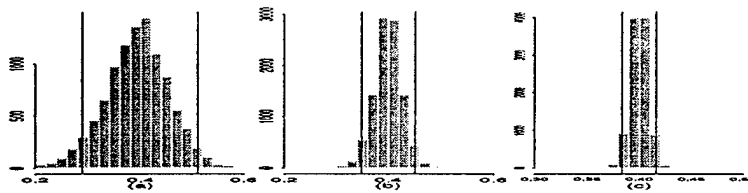


Figure 6: Histograms and theoretical 95% confidence intervals for Whittle estimate  $\{\bar{\phi}_1\}_n$  of  $\phi_1$  in model (3.17), (a)  $n = 200$ , (b)  $n = 1000$ , (c)  $n = 10,000$ .

## 5. Diagnostic checking

After identifying and estimating the parameters of times series model, it is always nice (although sometimes disconcerting) to see if the estimated model is really a good fit to the data. This is where diagnostic checking procedures come in. This usually involves identifying the residuals and seeing how well they match the distribution originally assumed for the innovations. In our case, this involves checking whether the residuals are i.i.d. under the assumption of an S $\alpha$ S distribution.

For infinite variance series we recommend the four following steps:

- (i) Graph the residuals; the pattern should follow a white noise model.
- (ii) Check that the ACF and PACF of the residuals are those of white noise.

The results and recommendations of Section 3 all apply here.

- (iii) The Durbin-Watson test of [PL];
- (iv) Various non-parametric tests, see [BD], p.312-313.

We ran a small simulation to see how well the second of these techniques actually worked. For models (3.17) and (3.18), stable parameters  $\alpha = 1.5, 1.75, 2$ ,

and sample sizes  $n = 200, 1000, 10,000$  we simulated data, identified a model using the techniques of Section 3, estimated parameters (using the Whittle estimator) and then checked whether the residuals were consistent with stable white noise model.

Table 10: Diagnostic checking based on ACF/PACF; models (3.17)–(3.18)					
Model	$n$	Ident.	$\hat{\beta}_n$	Diag. Check (C)	Diag. Check(G)
$\alpha = 1.5$					
(3.17)	200	AR(1) (C)	.355	WN	NOT WN
(3.18)	200	MA(1) (C) <sup>1</sup>	.867	WN	NOT WN
(3.17)	1000	AR(1)(C) <sup>2</sup>	.388	WN	WN
(3.18)	1000	MA(1) (C)	.785	WN	WN
(3.17)	10000	AR(1) (G)	.402	N/A	WN
(3.18)	10000	MA(1) (G)	.793	N/A	WN
$\alpha = 1.75$					
(3.17)	200	AR(1) (C)	.388	WN	WN
(3.18)	200	MA(1) (C) <sup>3</sup>	.835	WN	NOT WN
(3.17)	1000	AR(1) (C)	.400	WN	WN
(3.18)	1000	MA(1) (C) <sup>4</sup>	.820	WN	NOT WN
(3.17)	10000	AR(1) (G)	.404	N/A	WN
(3.18)	10000	MA(1) (G)	.797	N/A	NOT WN
$\alpha = 2.00$					
(3.17)	200	AR(1) (G)	.443	WN	WN
(3.18)	200	MA(1) (C) <sup>5</sup>	.767	WN	WN
(3.17)	1000	AR(1) (G)	.398	WN	WN
(3.18)	1000	MA(1) (C)	.822	WN	WN
(3.17)	10000	AR(1) (G)	.425	N/A	NOT WN
(3.18)	10000	MA(1) (G)	.801	N/A	WN

<sup>1</sup> Gaussian and true 1.5-stable limits indicate MA(11). AIC indicates MA(1).

<sup>2</sup> Cauchy limits indicate AR(1) or ARMA(1,1). AIC indicates AR(1).

<sup>3</sup> Gaussian and true 1.75-stable limits indicate MA(5). AIC indicates MA(1).

<sup>4</sup> Gaussian limits indicate MA(5). True 1.75-stable limits indicate possibly MA(8). AIC indicates MA(1) or MA(5)(the value is almost the same). For reasons of parsimony we chose MA(1).

<sup>5</sup> True Gaussian limits indicate MA(13), while Cauchy limits indicate correct model MA(1). AIC indicates MA(1).

Model identification was based on ACF/PACF analysis, unless more than one model seemed acceptable, in which case we differentiated between the models via the AIC. For the “small” sample sizes of  $n = 200, 1000$  we used Cauchy based confidence limits for the ACF/PACF analysis, while for  $n = 10,000$  we used Gaussian limits. In fitting the residuals, with sample sizes  $n = 200, 1000$

we used both Cauchy and Gaussian limits.

The results summarized in Table 10 show that diagnostic checking based on ACF/PACF works as well in stable case as in the Gaussian. The letters (C) or (G) in the table indicate that, respectively, Cauchy or Gaussian based confidence limits were used. The coefficient estimate  $\hat{\beta}_n$  is  $\bar{\phi}_1$  in the AR(1) (3.17) case, and  $\bar{\theta}_1$  in the MA(1) (3.18) case.

As a guide to reading the table, consider the first line, which indicates that an AR(1) series of length  $n = 200$  and stable innovations with  $\alpha = 1.5$  was generated according to the model (3.17). It was correctly identified as an AR(1) model, using Cauchy confidence intervals for the ACF and PACF. The estimate of the the AR coefficient was .355. (The actual value was 0.4; cf. (3.17).) The residuals were then investigated. Using Cauchy confidence intervals they were judged to be white, which was not the case with Gaussian confidence intervals.

Although we have not studied it, we note that a stable analogue of the Durbin-Watson statistic has been developed in [PL]. Since this statistic essentially checks that the ACF of the residual sequence at lag 1 is that of white noise, its asymptotic distribution is given by Theorem 3.1.

Unfortunately no other, stronger, tools (such as the  $\chi^2$  test in the finite variance case) are currently available for diagnostic checking in the stable situation. This would seem to be a promising and important direction for further research.

## 6. Acknowledgement

We are grateful to Dr. Aaron Gross who wrote the initial S-Plus programs that this study was based on.

## Appendix: McCulloch's quantile estimator of stable parameters

In this section we describe the estimator developed by McCulloch ([MC]) for the indices  $\alpha$  (of stability) and  $\beta$  (of skewness) of a stable distribution, which have been referred to in the body of the paper.

Let  $X \sim S_\alpha(\sigma, \beta, \mu)$  and denote the  $p$ -th quantile of this distribution by  $X_p$ . McCulloch's estimator uses five quantiles to estimate  $\alpha \in [0.6, 2.0]$  and  $\beta \in [-1, 1]$ , and is structured as follows:

Set

$$\Phi_1(\alpha, \beta) = \frac{X_{.95} - X_{.05}}{X_{.75} - X_{.25}},$$

$$\Phi_2(\alpha, \beta) = \frac{X_{.95} + X_{.05} - 2X_{.50}}{X_{.95} - X_{.05}}.$$

Since  $\Phi_1$  is monotonic in  $\alpha$  and  $\Phi_2$  is monotonic in  $\beta$  (for fixed  $\alpha$ ) we can we

can invert these functions to obtain

$$\alpha = \Psi_1(\Phi_1, \Phi_2),$$

$$\beta = \Psi_2(\Phi_1, \Phi_2).$$

McCulloch tabulated  $\Psi_1$  and  $\Psi_2$  for various values of  $\Phi_1$  and  $\Phi_2$ .

To form an estimator, take a random sample from a stable distribution and define  $\hat{\Phi}_1$  and  $\hat{\Phi}_2$  by replacing the quantiles  $X_k$  by the corresponding sample quantiles  $\hat{X}_k$ . Since the sample quantiles are consistent for the population quantiles,  $\hat{\Phi}_1$  and  $\hat{\Phi}_2$  are consistent estimators of  $\Phi_1$  and  $\Phi_2$ . Define

$$\hat{\alpha} = \Psi_1(\hat{\Phi}_1, \hat{\Phi}_2),$$

$$\hat{\beta} = \Psi_2(\hat{\Phi}_1, \hat{\Phi}_2).$$

Given a random sample from a stable distribution, we can now use the tables to find  $\hat{\alpha}$  and  $\hat{\beta}$ .

To obtain estimates of the scale and location parameters, McCulloch defined similar functions using these same 5 quantiles, which were also tabulated for various scale and location values. These tables can be used in a similar fashion so as to obtain estimates for  $\mu$  and for  $\sigma$ .

### References

- [BJ] G. E. P. Box and G. M. Jenkins *Time Series Analysis: Forecasting and Control* Wiley, New York, 1976.
- [BD] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer, New York, 1991.
- [BH] R. Bhansali, Consistent order determination for processes with infinite variance, *J. R. Statist. Soc. B*, Vol. 50 (1988), 46–60.
- [CD] M. Calder and R. A. Davis, Inference for linear processes with stable noise, *this volume*.
- [CH] G. Christoph, On some differences in limit theorems with a normal or a non-normal stable limit law, *Math. Nachr.*, Vol. 153 (1991), 247–256.
- [CMS] J. M. Chambers, C. L. Mallows and B. W. Stuck, A method for simulating stable random variables, *Journal of the American Statistical Association*, Vol. 71 (1976), 340–344.
- [DR1] R. A. Davis and S. Resnick, Limit theory for moving averages of random variables with regularly varying tail probabilities, *Ann. Prob.*, Vol. 13 (1985), 179–195.

- [DR2] R. A. Davis and S. Resnick, Limit theory for the sample covariance and correlation functions of moving averages, *Ann. Statist.*, Vol. 14 (1986), 533–558.
- [FE] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. II, 4th ed., Wiley, New York, 1983.
- [FR] P. D. Feigin and S. R. Resnick, Pitfalls of fitting autoregressive models for heavy-tailed time series, *in preparation*.
- [HA] P. Hall, Two-sided bounds on the rate of convergence to a stable law, *Z. Wahrsch. verw. Gebiete*, Vol. 57 (1981), 349–364.
- [HW] L. Heinrich and W. Wolf, On the convergence of  $U$ -statistics with stable limit distribution *J. Mult. An.*, Vol. 44 (1993), 266–278.
- [JM] A. Janssen and D. M. Mason, On the rate of convergence of sums of extremes to a stable law, *Probab. Theory and Rel. Fields*, Vol. 86 (1990), 253–264.
- [KN1] K. Knight, Consistency of Akaike's information criterion for infinite variance autoregressive processes, *Ann. Statist.*, Vol. 17 (1989), 824–840.
- [KN2] K. Knight, Rate of convergence of centred estimates of autoregressive parameters for infinite variance autoregressions, *J. Time Ser. Anal.*, Vol. 8 (1987), 51–60.
- [LPPR] R. LePage, K. Podgórski, M. Ryznar, Bootstrapping signs and permutations for regression with heavy tailed errors: a seamless resampling, *this volume*.
- [MC] J.H. McCulloch, Simple consistent estimators of stable distribution parameters, *Communications in Statistics – Computation and Simulation*, Vol. 15 (1986), 1109–1136.
- [MGKA] T. Mikosch, T. Gadirich, C. Klüppelberg and R.J. Adler, Parameter estimation for ARMA models with infinite variance innovations, *Annals of Statistics* Vol. 23 (1995), 305–326.
- [MI] T. Mikosch, Periodogram estimates from heavy tailed data, *this volume*.
- [NS] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*, Wiley, New York, 1995.
- [PDM] O. V. Pictet, M. M. Dacorogna, U. A. Müller, Behaviour of tail estimates under various distributions, *this volume*.

- [PL] P. C. B. Phillips, and M. Loretan, The Durbin-Watson ratio under infinite-variance errors, *J. of Econometrics*, Vol. 47 (1991), 85–114.
- [R1] S. Resnick, Heavy tail modelling and teletraffic data, *Ann. Statistics*, 1997.
- [R2] S. Resnick, Why non-linearities can ruin the heavy tailed modeller's day, *this volume*.
- [ST] G. Samorodnitsky, and M. Taqqu, *Stable Non-Gaussian Random Processes*, Chapman-Hall, New York, 1994.

Robert J. Adler  
Department of Statistics  
UNC Chapel Hill  
Chapel Hill, NC, 27599-3260  
and  
Faculty of Industrial Engineering  
Technion, Haifa, 32000, Israel.

adler@stat.unc.edu  
radler@ie.technion.ac.il

Raisa E. Feldman and Colin Gallagher  
Statistics and Applied Probability  
UC Santa Barbara  
Santa Barbara, CA, 93106-2050

epstein@pstat.ucsb.edu  
colin@pstat.ucsb.edu